弹性伸缩

用户指南

用户指南

使用须知

ECS实例生命周期的管理

加入伸缩组的 ECS 实例有两种类型：自动创建的 ECS 实例、手工添加的 ECS 实例。

自动创建的 ECS 实例

“自动创建的 ECS 实例”是指根据用户的伸缩配置和伸缩规则，由弹性伸缩服务自动创建的 ECS 实例。

弹性伸缩服务负责该类型 ECS 实例的全生命周期管理，即在伸缩组扩展活动中，负责该 ECS 实例的创建，在伸缩组的收缩活动中，负责停止和释放该 ECS 实例。

手工添加的 ECS 实例

手工添加的 ECS 实例 是指不是由弹性伸缩服务创建，但由用户手工添加到伸缩组中的 ECS 实例。

弹性伸缩服务不负责该类型 ECS 实例的全生命周期管理，即该类型的 ECS 实例是在弹性伸缩服务之外创建，只能手工加入伸缩组，在伸缩组的收缩活动中或用户手工移出该类型 ECS 实例时，弹性伸缩服务只负责将该 ECS 实例移出伸缩组，不会停止和释放该 ECS 实例。

实例状态

ECS 实例在伸缩组中的生命周期，通过以下几种状态描述：

- Pending – 表示 ECS 实例正在加入伸缩组，包括创建实例、加入负载均衡、添加 RDS 访问名单等过程。
- InService – 表示 ECS 实例已成功加入伸缩组，并正常提供服务。
- Removing – 表示 ECS 实例正在移出伸缩组。

实例的健康状态

ECS 实例在伸缩组中的健康状态为：

- Healthy(健康)
- Unhealthy(不健康)

本期仅通过 ECS 实例为非 **运行中（Running）** 状态来判断该实例不健康。弹性伸缩会自动移出伸缩组中不健康的 ECS 实例。

- 对于 **自动创建的 ECS 实例**，弹性伸缩会停止和释放该 ECS 实例。
- 对于 **手工添加的 ECS 实例**，弹性伸缩不会停止和释放该 ECS 实例。

冷却时间

冷却时间指成功执行伸缩活动后的一段锁定时间。您可以通过以下两种方式配置冷却时间：

在伸缩组中配置默认冷却时间，默认冷却时间不能为空，请参照 创建伸缩组。

在伸缩规则中配置冷却时间，如果不配置，将使用默认冷却时间，请参照 创建伸缩规则。

说明：如果同时在伸缩组和伸缩规则中配置了冷却时间，弹性伸缩服务将优先使用伸缩规则中的冷却时间。

冷却时间规则

伸缩组成功执行伸缩活动后，弹性伸缩服务开始计算冷却时间。如果在伸缩活动中，多台 ECS 实例加入或者移出伸缩组，则从最后一台 ECS 实例加入或者移出伸缩组后开始计算冷却时间，请参照 示例 1。如果在伸缩活动中，没有 ECS 实例成功加入或者移出伸缩组，则不会开始计算冷却时间。

在冷却时间内，伸缩组会拒绝由云监控报警任务触发的伸缩活动请求。但其他类型任务（手动执行任务、定时任务）触发的伸缩活动可以立即执行，绕过冷却时间。

如果您停用伸缩组，然后再次启用伸缩组，冷却时间将会失效，请参照 示例 2。

说明：冷却时间仅锁定同一个伸缩组的伸缩活动，不会对其他伸缩组的伸缩活动产生影响。

示例场景

示例 1

一个伸缩组 asg-uf6f3xewn3dvz4bsy7r1，默认冷却时间是 10 分钟，伸缩组内存在一条伸缩规则 add3，冷却时间为 15 分钟。

根据 add3 成功执行一次伸缩活动后，扩张了 3 台 ECS 实例，从第 3 台 ECS 实例加入伸缩组开始计算冷却时间，15 分钟内不会接受由云监控报警任务触发的伸缩活动请求。

示例 2

一个伸缩组 asg-m5efkz67re9x7a571bjh，默认冷却时间是 10 分钟，伸缩组内存在一条伸缩规则 remove1，没有设置冷却时间。

在 18:00 根据 remove1 成功执行一次伸缩活动后，收缩了 1 台 ECS 实例。正常情况下，在 18:10 前伸缩组不会接受由云监控报警任务触发的伸缩活动请求。此时停用伸缩组，然后在 18:05 再次启用伸缩组，冷却时间将失效。如果在 18:05 至 18:10 间云监控报警任务触发了伸缩活动请求，伸缩组仍会接受并执行请求。

伸缩组状态

伸缩组状态

伸缩组具有以下三种状态：Active、Inacitve 和 Deleting。

状 态	Open API 对应的状态
创建中	Inactive
已创建	Inactive
启用中	Inactive
运行中	Active
停用中	Inactive
已停止	Inactive
删除中	Deleting

伸缩活动的流程

伸缩活动的生命周期为：**判断伸缩组的健康状态和边界条件** 步骤与 **启动 cooldown** 步骤之间的所有活动。

伸缩组自动伸缩

自动扩展

判断伸缩组的健康状态和边界条件。

分配 ActivityId 和执行伸缩活动。

创建 ECS 实例。

修改 Total Capacity。

分配 IP。

添加 RDS 白名单。

启动 ECS 实例，等待启动完成。

挂载负载均衡，将权重设为创建伸缩配置时指定的 **负载均衡权重**。

伸缩活动完成，启动 cooldown。

自动收缩

判断伸缩组的健康状态和边界条件。

分配 ActivityId 和执行伸缩活动。

从负载均衡移除 ECS 实例。

停止 ECS 实例，等待停止完成。

从 RDS 白名单移除。

释放 ECS 实例。

修改 Total Capacity。

伸缩活动完成，启动 cooldown。

用户手工加入和移出既有的 ECS 实例

手工加入

判断伸缩组的健康状态、边界条件和 ECS 实例的状态、类型。

分配 ActivityId 和执行伸缩活动。

加入 ECS 实例。

修改 Total Capacity。

添加 RDS 白名单。

挂载负载均衡，将权重设为当前伸缩组中已激活的伸缩配置上指定的“负载均衡权重”。

注意：目前弹性伸缩要求手工加入实例时，实例的规格必须与伸缩组中已激活的伸缩配置上的规格相同，所以此处使用了伸缩配置上指定的“负载均衡权重”。

伸缩活动完成，启动 cooldown。

手工移出

判断伸缩组的健康状态和边界条件。

分配 ActivityId 和执行伸缩活动。

负载均衡停止向该 ECS 实例转发流量。

等待 60 秒后，从负载均衡移除 ECS 实例。

从 RDS 白名单移除。

修改 Total Capacity。

从伸缩组移出。

伸缩活动完成，启动 cooldown。

伸缩活动的状态

伸缩活动在请求阶段被拒绝，为 **Rejected** 状态。

伸缩活动执行过程中，为 **InProgress** 状态。

伸缩活动执行完成后，有 3 种状态：

成功（**Successful**）：根据 Minsize、MaxSize 调整后的伸缩规则，伸缩活动成功将所有 ECS 实例加入或移出伸缩组。

说明：加入伸缩组成功表明创建 ECS 实例、加入负载均衡实例、配置 RDS 访问白名单每一步骤都成功。以上任何步骤失败都认为是该 ECS 实例加入伸缩组失败。

部分成功（**Warning**）：根据 Minsize、MaxSize 调整后的伸缩规则，伸缩活动至少有一台ECS实例加入或移出伸缩组不成功。

全部失败（**Failed**）：根据 Minsize、MaxSize 调整后的伸缩规则，伸缩活动没有一台ECS实例成功加入或移出伸缩组。

举例

伸缩规则：伸缩规则定义增加 5 台 ECS 实例，当前伸缩组的 Total Capacity = 3，MaxSize = 5，则执行该伸缩规则时会调整成增加 2 台 ECS 实例。伸缩活动执行完成后，3 种状态对应的情况如下：

- 成功：2 台 ECS 实例全部创建成功，并正确加入负载均衡实例和配置 RDS 访问白名单。
- 部分成功：2 台 ECS 实例全部创建成功，但只有 1 台正确加入负载均衡实例和配置 RDS 访问白名单，另外一台因无法加入负载均衡实例和配置 RDS 访问白名单而被回滚操作释放。
- 全部失败：2 台 ECS 实例全部创建失败，或者 2 台 ECS 实例全部创建成功，但这 2 台 ECS 实例都因无法加入负载均衡实例和配置 RDS 访问白名单而被回滚操作释放。

伸缩活动中失败实例的回滚

伸缩活动有 ECS 实例加入伸缩组失败时，需要保持 ECS 实例级事务的完整性，而非伸缩活动级事务的完整性，即只进行 ECS 实例级回滚，而不是伸缩活动级回滚。

例如，当伸缩组创建了 20 台 ECS 实例，但只有 19 台 ECS 实例成功加入负载均衡时，则只对不成功的 1 台 ECS 实例进行自动释放操作。

由于弹性伸缩是借助阿里云的 RAM（Resource Access Management）服务，通过 ECS Open API代替用户弹性伸缩 ECS 实例资源，所以回滚的 ECS 实例仍然会被扣费。

移出不健康ECS实例

当 ECS 实例成功加入伸缩组后，弹性伸缩服务会定期扫描该 ECS 实例的运行状态，如果发现该 ECS 实例为非**运行中**（Running）状态，则会将该 ECS 实例移出伸缩组。

- 对于弹性服务自动创建的 ECS 实例：将立即启动移出和释放该 ECS 实例操作。
- 对于用户手工添加的 ECS 实例：将立即启动移出该 ECS 实例操作，但不会停止和释放该 ECS 实例。

移出不健康 ECS 实例，不受 MinSize 的限制，即移出不健康 ECS 实例后，有可能导致 Total Capacity 低于 MinSize，此时弹性伸缩服务会自动创建差额的 ECS 实例使得 Total Capacity 等于 MinSize。

通知

需要短信息和邮件通知的伸缩活动条件如下：

- 只有定时任务、云监控报警任务、健康检查触发的伸缩活动才通知。
- 只有创建了 ECS 实例或释放了 ECS 实例才通知。

如满足以上条件，则一个伸缩活动对应一条通知。

强制干预

弹性伸缩服务不会阻止用户从 ECS 控制台删除自动创建的 ECS 实例等强制干预行为。对于用户的强制干预，弹性伸缩服务的处理方法如下：

资源	强制干预类型	弹性伸缩服务的处理方法
ECS	通过 ECS 控制台或 OpenAPI 强制删除伸缩组里的 ECS 实例	会按照健康检查来判断该 ECS 实例为不健康状态，并将该 ECS 实例移出伸缩组。（该 ECS 实例在 RDS 实例访问白名单中的内网 IP 无法自动删除。）如移出该 ECS 实例后，使得 Total Capacity 低于 MinSize，弹性伸缩会自动创建差额的 ECS 实例使得 Total Capacity 等于 MinSize。
ECS	客户取消将 ECS Open API 权限授权予弹性伸缩	弹性伸缩将拒绝所有伸缩活动请求。
负载均衡	通过负载均衡控制台或 OpenAPI 将 ECS 实例强制移出负载均衡实例	弹性伸缩不会自动感知用户的移出行为也不会处理该类异常状态，该 ECS 实例仍存在伸缩组中。当发生收缩活动时，如果按照移出策略挑选到该 ECS 实例，仍然会把该 ECS 实例释放掉。
负载均衡	通过负载均衡控制台或 OpenAPI 强制删除负载均衡实例或关闭负载均衡实例的健康检查。	对于配置该负载均衡实例的伸缩组，不会再往伸缩组加入 ECS 实例。伸缩任务可以触发伸缩规则从伸缩组移出 ECS 实例，健康检查判断某个 ECS 实例不健康，也可将该 ECS 实例移出伸缩组。
负载均衡	系统导致负载均衡实例不可用（如欠费、故障）	除了用户手工触发的移出 ECS 实例的伸缩活动，其它所有伸缩活动执行失败。
负载均衡	客户取消将负载均衡 Open API权限授权予弹性伸缩。	对于配置了负载均衡实例的伸缩组，弹性伸缩将拒绝所有伸缩活动请求。
RDS	通过RDS控制台或OpenAPI将 ECS 实例的ip强制移出 RDS 访问白名单	弹性伸缩不会自动感知用户的移出行为也不会处理该类异常状态，该 ECS 实例仍存在伸缩组中。当发生收缩活动时，如果按照移出策略挑选到该 ECS 实例，仍然会把该 ECS 实例释放掉。
RDS	通过 RDS 控制台或 OpenAPI 强制删除 RDS 实例	对于配置该 RDS 实例的伸缩组，不会再往伸缩组加入 ECS 实例。伸缩任务可以触发伸缩规则

		从伸缩组移出 ECS 实例，健康检查判断某个 ECS 实例不健康，也可将该 ECS 实例移出伸缩组。
RDS	系统导致 RDS 实例不可用（如欠费、故障）	除了用户手工触发的移出 ECS 实例的伸缩活动，其它所有伸缩活动执行失败。
RDS	客户取消将 RDS Open API 权限授权予弹性伸缩。	对于配置了 RDS 实例的伸缩组，弹性伸缩将拒绝所有伸缩活动请求。

数量限制

目前，弹性伸缩服务的功能项存在以下数量限制：

一个账号下最多创建 20 个伸缩组。

- 一个伸缩组内最多创建 10 个伸缩配置。
- 一个伸缩组内最多创建 50 个伸缩规则。
- 一个伸缩组内最多创建 6 个事件通知。
- 一个伸缩组内最多创建 6 个生命周期挂钩。

对于所有地域和所有伸缩组，一个账号下最多弹性伸缩 1000 台 ECS 实例。此数量只包含自动创建的 ECS 实例，不包含手动添加的 ECS 实例。

一个账号下最多创建 20 个定时任务。

注意事项

伸缩规则

在计算和执行过程中，伸缩规则可以根据伸缩组的 MinSize、MaxSize 进行自动调整其需要增加或减少的 ECS 实例数（例：如伸缩规则中指定将伸缩组的 ECS 实例数调整至 50 台，但伸缩组 MaxSize 只有 45 台，则整个伸缩规则会按调整至 45 台来计算和执行。）

伸缩活动

同一伸缩组内、同一时刻只能有一个伸缩活动在执行。

伸缩活动不可以中断。例如，某个创建 20 台 ECS 实例的伸缩活动正在执行中，当创建到第 5 台 ECS 实例时，您无法强行终止该伸缩活动。

伸缩活动有 ECS 实例加入伸缩组失败时，需要保持 ECS 实例级事务的完整性，而非伸缩活动级事务的完整性，即只进行 ECS 实例级回滚，而不是伸缩活动级回滚。例如，当伸缩组创建了 20 台 ECS 实例，但只有 19 台 ECS 实例成功加入负载均衡时，则只对不成功的 1 台 ECS 实例进行自动释放操作。

由于弹性伸缩是借助阿里云的 RAM （Resource Access Management）服务，通过 ECS Open API代替用户弹性伸缩 ECS 实例资源，所以回滚的 ECS 实例仍然会被扣费。

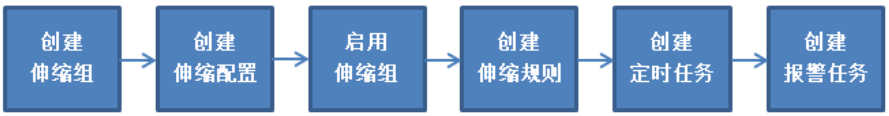
冷却时间

在冷却时间内，伸缩组只会拒绝云监控报警任务类型的伸缩活动请求，其他类型的触发任务（如用户手工执行伸缩规则、定时任务等）可以绕过冷却时间立即执行伸缩活动。

每个伸缩活动的最后一个 ECS 实例加入或移出伸缩组成功后，整个伸缩组冷却时间才开始计时。

使用流程

创建完整的弹性伸缩方案，您需要通过以下步骤来完成：

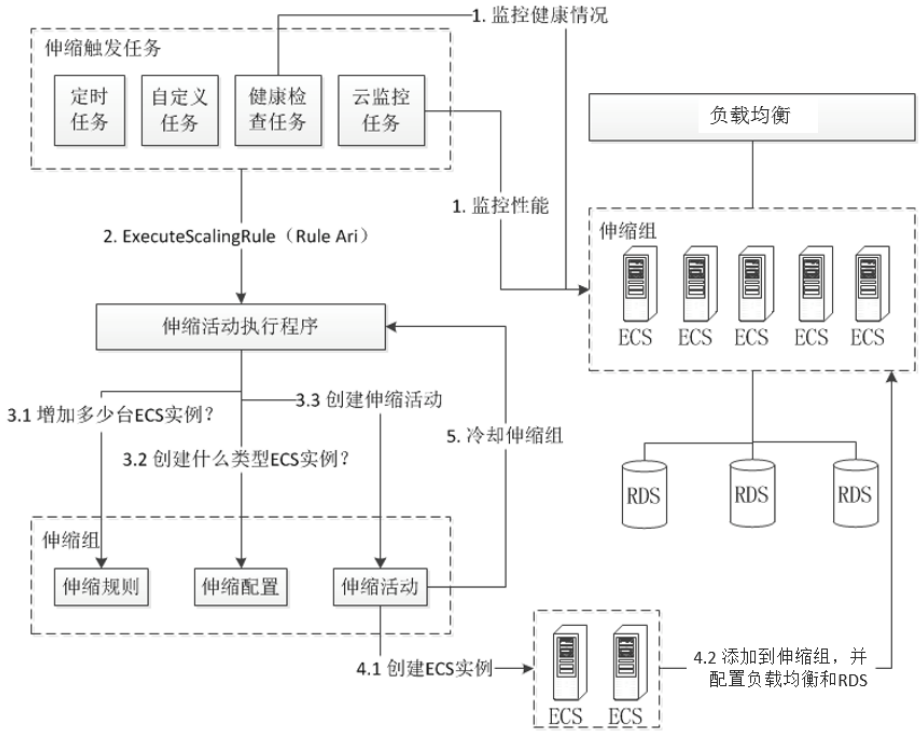


1. 创建伸缩组（CreateScalingGroup），配置伸缩资源的最小值（MinSize）、最大值（MaxSize）及需要关联的负载均衡实例和RDS实例。
2. 创建伸缩配置（CreateScalingConfiguration），指定需要弹性伸缩的 ECS 实例的相关属性，如 ImageID、InstanceType 等。

3. 以第二步创建的伸缩配置启用伸缩组（EnableScalingGroup）。
4. 创建伸缩规则（CreateScalingRule），如 **加 N 台 ECS 实例** 的伸缩规则。
5. 创建定时任务（CreateScheduledTask），如创建 12：00 触发第四步伸缩规则的定时任务。
6. 创建报警任务（云监控 API PutAlarmRule），如创建 CPU 平均值 （也可以是最大值或最小值）大于等于 80% 则增加一台 ECS 实例的报警任务。

工作流程

下图是弹性伸缩的工作流程。



创建好伸缩组、伸缩配置、伸缩规则、伸缩触发任务后，系统会自动化执行以下流程（以增加 ECS 实例为例）：

伸缩触发任务会按照各自 **触发生效的条件** 来触发伸缩活动。

- 云监控任务会实时监控伸缩组内 ECS 实例的性能，并根据用户配置的报警规则（如伸缩组内所有 ECS 实例的 CPU 平均值大于 60%）触发执行伸缩规则请求。
- 定时任务会根据用户配置的时间来触发执行伸缩规则请求。
- 您可以根据自己的监控系统及相应的报警规则（如在线人数、作业队列）来触发执行伸缩规则请求。
- 健康检查任务会定期检查伸缩组和 ECS 实例的健康情况，如发现有不健康的 ECS 实例（如 ECS 为非running 状态）会触发执行 **移出该 ECS 实例** 的请求。

系统自动通过 ExecuteScalingRule 接口触发伸缩活动，并在该接口中指定需要执行的伸缩规则的阿里云资源唯一标识符（Ari）。

如果是用户自定义的任务，则需要用户在自己的程序中调用 ExecuteScalingRule 接口来实现。

根据步骤 2 传入的伸缩规则 Ari（Rule Ari）获取伸缩规则、伸缩组、伸缩配置的相关信息，并创建伸缩活动。

i. 通过伸缩规则 Ari 查询伸缩规则以及相应的伸缩组信息，计算出需要增加的 ECS 实例数量，并获得需要配置的负载均衡和 RDS 信息。
ii. 通过伸缩组查询到相应的伸缩配置信息，即获得了需要创建的ECS实例的配置信息（CPU、内存、带宽等）。
iii. 根据需要增加的 ECS 实例数量、ECS 实例配置信息、需要配置的负载均衡实例和 RDS 实例创建伸缩活动。

在伸缩活动中，自动创建 ECS 实例并配置负载均衡和 RDS。

i. 按照实例配置信息创建指定数量的 ECS 实例。
ii. 将创建好的 ECS 实例的内网 IP 添加到指定的 RDS 实例的访问白名单当中，将创建好的 ECS 实例添加到指定的负载均衡实例当中。

伸缩活动完成后，启动伸缩组的冷却功能。待冷却时间完成后，该伸缩组才能接收新的执行伸缩规则请求。

管理 ECS 实例模板

ECS 实例模板概述

在您的业务需求增长时，弹性伸缩可以根据伸缩规则进行弹性扩张，自动创建 ECS 实例并添加到伸缩组中。

伸缩配置是弹性伸缩用于创建 ECS 实例的模板。在创建伸缩配置时，您可以指定 ECS 实例的信息，例如实例规格、镜像类型、存储大小、登录实例用的密钥对等。您也可以修改现有伸缩配置，快速满足新的业务需求。

说明：伸缩配置依托伸缩组，因此要确保已经存在一个伸缩组，关于如何创建伸缩组，请参阅 创建伸缩组。一个伸缩组内可以创建的伸缩配置数量有限，请参阅 使用限制。

目前，您可以进行以下伸缩配置相关的操作：

- 创建伸缩配置
- 选用伸缩配置
- 修改伸缩配置
- 删除伸缩配置

创建伸缩配置

创建伸缩配置的流程和创建 ECS 实例类似，但是由于应用场景为弹性伸缩，因此不支持配置某些项目，例如地域，实际操作时请以控制台页面为准。页面上的各配置项均配有简洁说明，如果您想深入了解各配置项的含义和用法，请参阅 使用向导创建实例。

操作步骤

请参照以下步骤创建一个伸缩配置：

登录弹性伸缩控制台，在伸缩组的 **操作** 列下单击 **管理**。

前往 **伸缩配置** 页面，单击 **创建伸缩配置** 按钮。

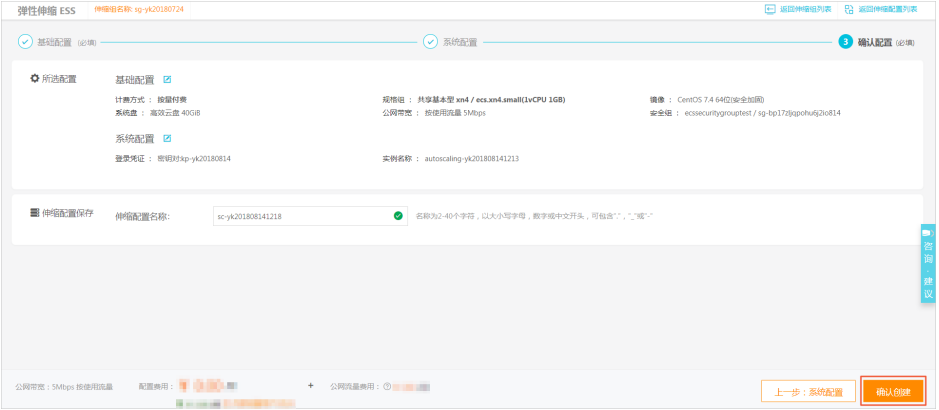
在 **基础配置** 页面中，依次配置计费方式、实例、镜像、存储、公网带宽和安全组，然后单击 **下一步：系统配置** 按钮。

说明：弹性伸缩仅支持创建 **按量付费实例** 和 **抢占式实例**，仅支持选择系统镜像、自定义镜像和共享镜像。

在 **系统配置** 页面中，依次配置标签（可选）、登录凭证、实例名称（可选）和高级选项（可选），然后单击 **下一步：确认配置** 按钮。

说明：高级选项包括实例 RAM 角色和实例自定义数据，选择专有网络的伸缩组方可配置高级选项。

在 **确认配置** 页面中，检查所选配置，填写伸缩配置名称，然后单击 **确认创建** 按钮。



在 **创建成功** 对话框中，单击 **启用配置** 按钮启用该伸缩配置。如果您暂时不需要使用该伸缩配置，可以直接关闭对话框。

创建成功后，伸缩配置会出现在伸缩配置列表中，您可以继续进行查看伸缩配置详情、选用伸缩配置等操作。

选用伸缩配置

在伸缩组中，伸缩配置有两种可能状态：生效 和 失效。您可以根据业务需要选用需要生效的伸缩配置，伸缩组只会根据生效的伸缩配置自动创建 ECS 实例。

伸缩配置内容较多，您可能会忘记具体的配置项。因此弹性伸缩还提供了查看伸缩配置详情的功能，供您随时了解各个伸缩配置，灵活选用 ECS 实例的模板。

操作步骤

请参照以下步骤查看一个伸缩配置的详情并选用一个伸缩配置：

登录弹性伸缩控制台，在伸缩组的 **操作** 列下单击 **管理**。

前往 **伸缩配置** 页面，在指定伸缩配置的 **操作** 列下单击 **查看详情**。

确定伸缩配置详情无误后，单击 **选用** 启用该伸缩配置。

说明：选用一个伸缩配置后，其他伸缩配置将被置于 失效 状态。

选用一个伸缩配置后，在满足弹性扩张条件时，对应的伸缩组会根据该伸缩配置自动创建 ECS 实例。

修改伸缩配置

如果现有伸缩配置都无法完全满足您的业务需求，但是存在一个和期望伸缩配置大体相同的伸缩配置，您可以修改该伸缩配置，调整特定配置项，无须重新创建一个伸缩配置，提高运维效率。

操作步骤

请参照以下步骤修改一个伸缩配置：

登录弹性伸缩控制台，在伸缩组的 **操作** 列下单击 **管理**。

前往 **伸缩配置** 页面，在指定伸缩配置的 **操作** 列下单击 **修改**。

在 **基础配置** 页面中，修改配置项，然后单击 **下一步：系统配置** 按钮。

在 **系统配置** 页面中，修改配置项，然后单击 **下一步：确认配置** 按钮。

在 **修改成功** 对话框中，单击 **启用配置** 按钮启用该伸缩配置。如果您暂时不需要使用该伸缩配置，可以直接关闭对话框。

修改一个伸缩配置后，之前根据该伸缩配置创建的 ECS 实例仍然可以正常运行。

删除伸缩配置

如果您不再需要一个伸缩配置，可以将其删除，空出伸缩配置的配额，用于创建其他伸缩配置。

前提条件

在删除一个伸缩配置前，请确保满足以下条件，否则删除操作会失败：

- 该伸缩配置必须处于 失效 状态。
- 对应的伸缩组中不存在根据该伸缩配置自动创建出的 ECS 实例。

操作步骤

请参照以下步骤删除一个或多个伸缩配置：

登录弹性伸缩控制台，在伸缩组的 **操作** 列下单击 **管理**。

前往 **伸缩配置** 页面，在指定伸缩配置的 **操作** 列下单击 **删除**。

您也可以勾选伸缩配置左侧的复选框，然后单击下方的 **删除** 按钮进行批量删除。

在 **删除伸缩配置** 对话框中，单击 **确认** 按钮。

管理单个伸缩组

伸缩组管理概述

本章节介绍与控制单个伸缩组有关的功能，包括伸缩组、伸缩规则和生命周期挂钩，这三项功能为进行伸缩活动提供了操作支持，您可以灵活控制伸缩组的活动，满足业务需要。

> **说明**：弹性伸缩服务支持的伸缩组、伸缩规则和生命周期挂钩数量有限，请参阅 使用限制。

假设出于业务需要，您需要在一段时间内使用弹性伸缩服务，实现按需增减 1 至 10 台 ECS 实例，并且在启动 ECS 实例前安装软件。

接下来将针对该场景，为您逐步展示通过弹性伸缩满足需求的过程，涉及到的操作有：

- 创建伸缩组：创建一个伸缩组，用于承载伸缩活动。
- 创建伸缩规则：创建一条伸缩规则，用于定义伸缩活动动作。
- 执行伸缩规则：执行一条伸缩规则，用于启动伸缩活动。
- 创建生命周期挂钩：创建一个生命周期挂钩，用于在伸缩活动前保留实例操作时间。
- 修改生命周期挂钩：修改一个生命周期挂钩，用于修改挂钩属性，适应新场景。
- 删除生命周期挂钩：删除一个不再使用的生命周期挂钩。
- 修改伸缩规则：修改一条伸缩规则，用于修改伸缩规则属性，适应新场景。
- 删除伸缩规则：删除一条不再使用的伸缩规则。

- 更改伸缩组状态：用于控制伸缩组启用/停用状态。
- 修改伸缩组：修改一个伸缩组，用于修改伸缩组属性，适应新场景。
- 删除伸缩组：删除一个不再使用的伸缩组。

实现自动伸缩

创建伸缩组

如果想使用弹性伸缩调整弹性计算资源大小，您首先需要拥有一个伸缩组。

什么是伸缩组

伸缩组是具有相同应用场景的 ECS 实例的集合。您可以在伸缩组内定义最大实例数、最小实例数和冷却时间等属性，还可以为这些 ECS 实例连接 SLB 和 RDS 实例，便于进行综合管理。

说明：一个账号下可以创建的伸缩组数量有限，请参阅 使用限制。

操作步骤

请参照以下步骤创建一个伸缩组：

登录弹性伸缩控制台。

在 **伸缩组管理** 页面中，单击 **创建伸缩组**。

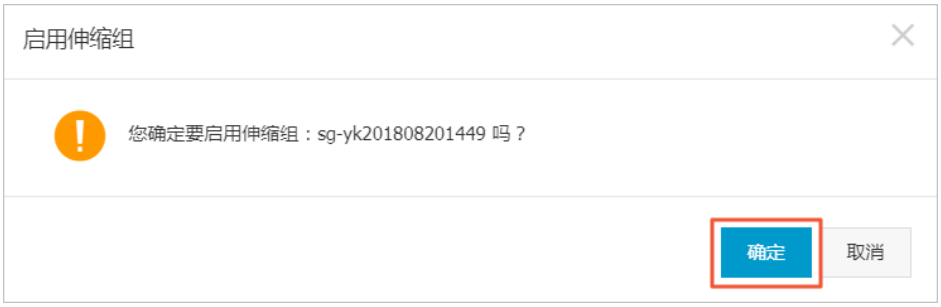
在 **创建伸缩组** 对话框中，依次指定伸缩组名称、伸缩最大实例数、伸缩最小实例数、默认冷却时间、移出策略、网络类型、负载均衡（可选）和数据库（可选），然后单击 **提交**。

说明：有关伸缩组的属性，请参阅 **伸缩组属性**。

在 **伸缩组创建成功** 对话框中，单击 **创建伸缩配置** 为伸缩组添加一个伸缩配置。

说明：有关伸缩配置，请参阅 **ECS 实例模板概述**。

在 **启用伸缩组** 对话框中，单击 **确定**。



伸缩组属性

以下表格描述了具体的伸缩组属性含义和示例。

名称	说明	示例
伸缩组名称	名称为2-40个字符，以大小写字母、数字或中文开头，可包含英文句号（.）、下划线（_）或短横线（-）。	sg-yk201808201449
伸缩最大实例数（台）	伸缩组内 ECS 实例数上限。超过上限时，弹性伸缩会根据移出策略自动移出 ECS 实例，使得伸缩组的当前 ECS 实例数等于上限。	10
伸缩最小实例数（台）	伸缩组内 ECS 实例数下限。低于下限时，弹性伸缩会自动添加 ECS 实例，使得伸缩组的当前 ECS 实例数等于下限。	1
默认冷却时间（秒）	伸缩组发生伸缩活动后的默认冷却时间。更多信息，请参阅 冷却时间。	600
移出策略	伸缩组内 ECS 实例数超过上限时，移出 ECS 实例的策略。更多信息，请参阅 移出策略。	最早伸缩配置对应的实例
网络类型	伸缩组所在的网络类型。您可以为专有网络伸缩组配置多台虚拟交换机，实现多可用区扩缩容，当无法在一个虚拟交换机所在的可用区内创建 ECS 实例时，弹性伸缩自动切换至其它可用区创建 ECS 实例。	经典网络
负载均衡	需要拥有 SLB 实例，请参阅 负载均衡 SLB。	slb-yk201807061512
数据库	需要拥有 RDS 实例，请参阅 云数据库 RDS。	5.7基础版

说明：伸缩组、SLB 实例和 RDS 实例必须位于同一个地域。

创建伸缩规则

在 创建伸缩组 中，您已经成功启用了一个伸缩组。但如果想弹性扩张或收缩 ECS 资源，您还需要创建一定的伸缩规则。

什么是伸缩规则

伸缩规则指弹性伸缩服务弹性扩张或收缩 ECS 资源时所依据的规则，目前支持以下三种规则：

- 调整至 N 台：执行伸缩规则后，服务中的实例数将被调整至 N 台。
- 增加 N 台：执行伸缩规则后，服务中的实例数在当前数量基础上增加 N 台。
- 减少 N 台：执行伸缩规则后，服务中的实例数在当前数量基础上减少 N 台。

说明：一个伸缩组内可以创建的伸缩规则数量有限，请参阅 使用限制。

如果执行伸缩规则后，伸缩组内服务中实例数将大于 *伸缩最大实例数* 或者小于 *伸缩最小实例数*，弹性伸缩服务会自动调整扩张或收缩的 ECS 实例数量，保证弹性伸缩结果不会超过数量限制。

示例

您拥有一个伸缩组 asg-bp19ik2u5w7esjcucu28，伸缩最大实例数为 3，伸缩规则 add3 为增加 3 台实例。如果当前服务中实例数为 2，执行伸缩规则 add3 时，仅会增加 1 台 ECS 实例。

您拥有一个伸缩组 asg-bp19ik2u5w7esjcucu28，伸缩最小实例数为 2，伸缩规则 reduce2 为减少 2 台实例。如果当前服务中实例数为 3，执行伸缩规则 reduce2 时，仅会减少 1 台 ECS 实例。

操作步骤

请参照以下步骤创建伸缩规则：

在 **伸缩组管理** 页面中，单击要创建伸缩规则的伸缩组 **操作** 列下的 **管理**。

前往 **伸缩规则** 页面，单击 **创建伸缩规则**。

在 **创建伸缩规则** 对话框中，依次指定规则名称、规则和冷却时间，然后单击 **创建伸缩规则** 按钮。

说明：冷却时间为可选项，如果不填写，则默认为伸缩组的冷却时间。

执行伸缩规则

在 创建伸缩规则 中，您已经成功创建了伸缩规则，接下来就可以执行一条伸缩规则，实现弹性扩张或收缩 ECS 资源。

限制条件

如果您需要执行伸缩规则，请注意以下限制条件：

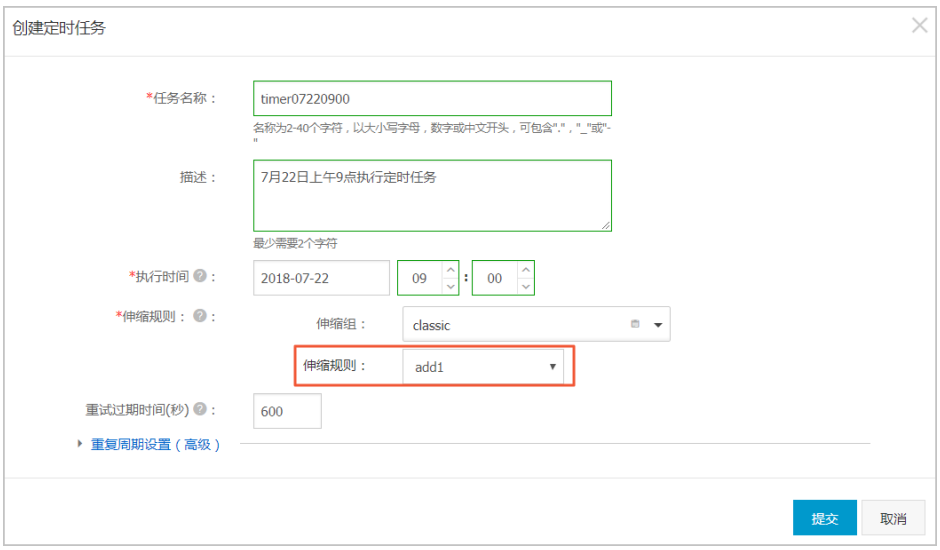
- 伸缩规则所在的伸缩组的状态必须为启用。
- 伸缩规则所在的伸缩组内不能存在执行中的伸缩活动。
- 对于所有地域和所有伸缩组，一个账号可以弹性伸缩的 ECS 实例数量有限，请参阅 使用限制。

目前，您可以通过三种方式执行一条伸缩规则：

- 通过定时任务执行
- 通过报警任务执行
- 手动执行

通过定时任务执行

在创建定时任务时，选择一条伸缩规则。在指定的时间点，弹性伸缩服务将自动执行该伸缩规则。



通过报警任务执行

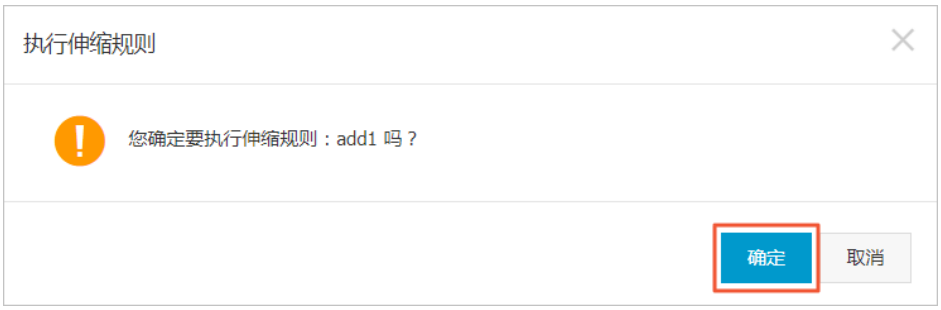
在创建报警任务时，选择一条报警触发规则。在满足报警条件时，弹性伸缩服务将自动执行该伸缩规则。

手动执行

当伸缩组内不存在执行中的伸缩活动时，手动执行可以绕过冷却时间。请参照以下步骤手动执行一条伸缩规则：

在 **伸缩规则** 页面中，在伸缩规则的 **操作** 列下单击 **执行**。

在 **执行伸缩规则** 对话框中，单击 **确定**。



如果执行成功，页面右上角会出现成功提示。

如果执行失败，会弹出错误提示。

您可以前往 **伸缩活动** 页面，查看具体的成功或失败信息。

创建生命周期挂钩

在 执行伸缩规则 中，您已经成功实现弹性扩张或收缩 ECS 资源。但是，这些 ECS 资源只进行了简单的初始配置，面对复杂的业务情况，您可能希望在启用这些 ECS 资源前进行自定义操作，这时就需要使用生命周期挂钩。

什么是生命周期挂钩

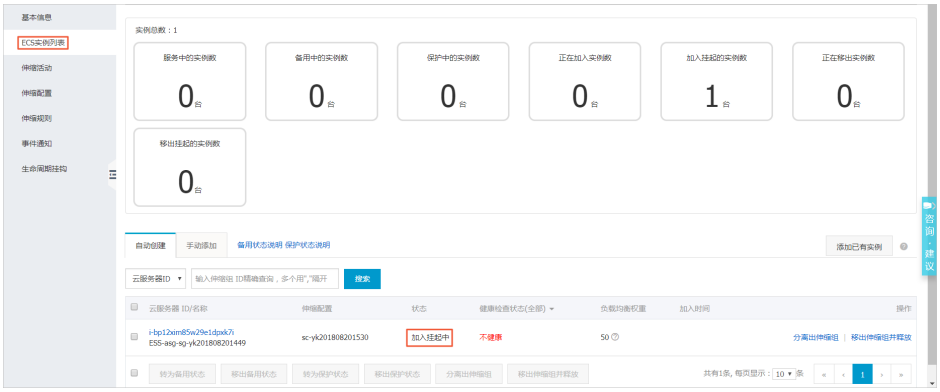
如果您为伸缩组创建了生命周期挂钩，在伸缩组进行伸缩活动时，正在加入或正在移出伸缩组的实例将被挂钩挂起并置于等待状态。生命周期挂钩仅在自动创建或移出 ECS 实例时生效，手动添加或移出 ECS 实例时不受其影响。

> **说明**：一个伸缩组内可以创建的生命周期挂钩数量有限，请参阅 使用限制。

示例

您拥有一个伸缩组 sg-yk201808201449，伸缩最小实例数为 0，伸缩组内有 1 个适用于弹性扩张活动的生命周期挂钩，目前伸缩组内没有 ECS 实例。

将伸缩最小实例数修改为 1，由于 ECS 实例数没有达到下限，弹性扩张活动被触发，伸缩组自动创建一台 ECS 实例。但是，由于存在生命周期挂钩，该 ECS 实例不会马上进入服务中状态，而是进入加入挂起中状态。



在超时时间结束前，您可以登录该 ECS 实例，进行预装软件或其他自定义操作。

功能特性

在实例保持等待状态的时间内，当前伸缩组将具有以下特性：

- 伸缩组不再运行其它的伸缩活动。
- 为您保留指定时长的操作时间（即挂钩的超时时间，请参阅 生命周期挂钩属性），您可以在挂起期间执行自定义操作，例如，初始化 ECS 实例配置或者获取 ECS 实例数据。
- 您可以删除生命周期挂钩来恢复执行伸缩活动。
- 您可以调用相关 OpenAPI 来结束生命周期活动或者删除生命周期挂钩。

操作步骤

请参照以下步骤创建一个生命周期挂钩：

前往 **生命周期挂钩** 页面，单击 **创建生命周期挂钩**。

在 **创建生命周期挂钩** 对话框中，依次指定名称、适用的伸缩活动类型、超时时间、执行策略、通知方式、MNS 主题和通知标识，然后单击 **创建生命周期挂钩**。

说明：有关生命周期挂钩的属性，请参阅 生命周期挂钩属性。

生命周期挂钩属性

以下表格描述了具体的生命周期挂钩属性含义和示例。

名称	说明	示例

名称	生命周期挂钩名称（2-40个字符），以大小写字母、数字或中文开头，可包含英文句号（.）、下划线（_）或短横线（-）。挂钩名字一旦确定，不能修改。	hz_yk0626
适用的伸缩活动类型	适用的伸缩活动类型。	弹性收缩活动
超时时间	默认情况下，实例保持等待状态的时间。最小为 30 秒，最大为 21600 秒（6 小时），必须为整数。	600
执行策略	执行策略包括 **继续** 和 **拒绝**。 - **继续**：继续执行弹性扩张活动或者弹性收缩活动。 - **拒绝**：针对弹性扩张活动，会直接释放创建出来的 ECS 实例；针对弹性收缩活动，无影响。	继续
通知方式	通知方式包括 **MNS 主题** 和 **MNS 队列**。选择通知方式后，您还需要选择具体的 MNS 主题或者 MNS 队列，如果尚未创建 MNS 主题或者 MNS 队列，请 **创建主题** 或者 **创建队列**。	MNS 主题
通知标识	弹性伸缩每次推送消息到通知对象时，会同时发送您预先指定的通知标识，便于管理和标记不同类别的通知信息。	General information

移出策略

有两种移出策略：默认策略和自定义策略。

默认移出策略

先按最早伸缩配置(OldestScalingConfiguration)对应的 ECS 实例进行一级筛选，再按最早创建的 ECS 实例(OldestInstance)进行二级筛选。

先在伸缩组里选最早伸缩配置(OldestScalingConfiguration)对应的 ECS 实例，然后在筛选的结果中，选最早创建的 ECS 实例(OldestInstance)。如果还有多台 ECS 实例，则随机选择一台移出伸缩组。

由于用户手工加入的既有 ECS 实例不与伸缩配置绑定，所以不会先选出该类型的 ECS 实例。

如果伸缩配置关联的 ECS 实例已经全部移出，但还需移出 ECS 实例时，则选最早用户手工移入的 ECS 实例。

自定义设置释放策略

根据用户设置的多条策略依次筛选 ECS 实例并移出伸缩组。

释放策略类型

OldestInstance：取最早创建的 ECS 实例，如果作为一级筛选条件，不区分是用户手工加入的 ECS 实例还是自动创建的 ECS 实例，直接按最早创建的实例进行移出。

NewestInstance：取最新创建的实例，如果作为一级筛选条件，不区分是用户手工加入的 ECS 实例还是自动创建的 ECS 实例，直接按最新创建的 ECS 实例进行移出。

OldestScalingConfiguration：取最早伸缩配置创建的 ECS 实例；对于用户手工移入的 ECS 实例，自动跳过此条件；如果与伸缩配置相关联的 ECS 实例已经全部移出，还要继续移出 ECS 实例时，则在用户手工加入的 ECS 实例中（没有与伸缩配置关联的 ECS 实例），随机选一台。

维护自动伸缩

修改生命周期挂钩

创建生命周期挂钩 后，如果当前挂钩属性不再满足需求，您可以直接修改一项或者多项属性，无需重新创建。

操作步骤

请参照以下步骤修改一个生命周期挂钩的属性：

在 **生命周期挂钩** 页面中，单击要修改的挂钩 **操作** 列下的 **修改**。

在 **修改生命周期挂钩** 对话框中，根据需要修改属性，然后单击 **修改生命周期挂钩**。

说明：有关生命周期挂钩的属性，请参阅 创建生命周期挂钩 下的 **生命周期挂钩属性**。不能修改挂钩名称，其他属性均可修改。

删除生命周期挂钩

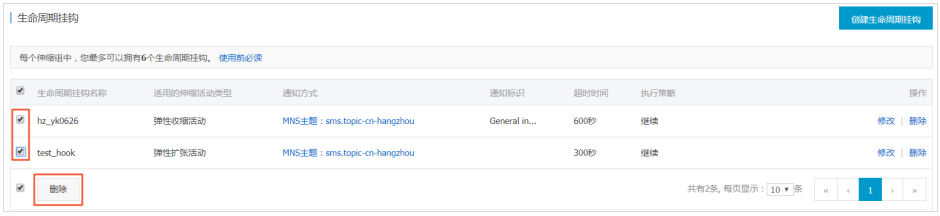
如果不再需要一个生命周期挂钩，您可以将其删除，空出挂钩配额。

操作步骤

请参照以下步骤删除一个生命周期挂钩：

在 **生命周期挂钩** 页面中，单击要删除的生命周期挂钩 **操作** 列下的 **删除**。

您也可以勾选生命周期挂钩左侧的复选框，然后单击下方的 **删除** 按钮进行批量删除。



在 **删除生命周期挂钩** 对话框中，单击 **确定** 完成删除操作。

说明：如果生命周期挂钩已经挂起一个 ECS 实例，在删除生命周期挂钩时，该 ECS 实例的等待状态会提前结束。

修改伸缩规则

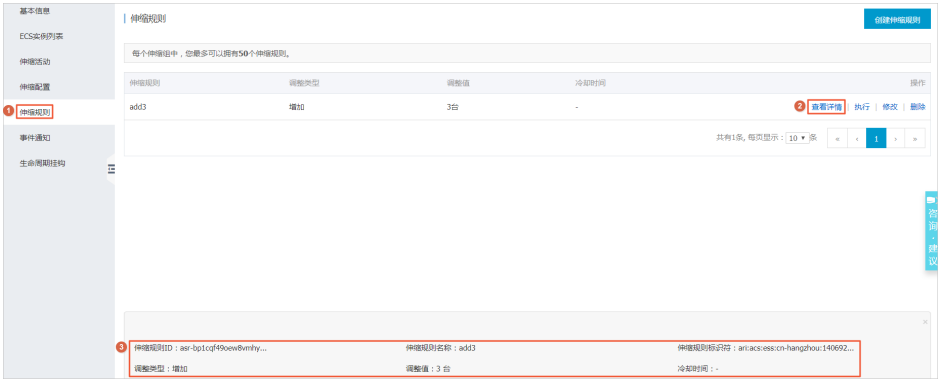
创建伸缩规则 后，如果当前伸缩规则配置不再满足需求，您可以直接修改一项或者多项属性，无需重新创建。

在修改前，您也可以查看伸缩规则的详情。

操作步骤

请参照以下步骤修改一条伸缩规则的属性：

前往 **伸缩规则** 页面，单击要修改的伸缩规则 **操作** 列下的 **查看详情**。



确定需要修改该伸缩规则后，单击 **修改**。

在 **修改伸缩规则** 对话框中，根据需要修改属性，然后单击 **修改伸缩规则**。

说明：有关伸缩规则的属性，请参阅 **创建伸缩规则** 下的 **什么是伸缩规则**。

删除伸缩规则

如果不再需要一条伸缩规则，您可以将其删除，空出伸缩规则配额。

操作步骤

在 **伸缩规则** 页面中，单击要删除的伸缩规则 **操作** 列下的 **删除**。

在 **删除伸缩规则** 对话框中，单击 **确定**。

更改伸缩组状态

创建伸缩组后，您可以控制伸缩组的状态。如果暂时不需要使用一个伸缩组，停用伸缩组即可，在需要使用时再启用。

操作步骤

在伸缩组的详情页面中，单击 <。

在 **伸缩组管理** 页面中，单击待停用伸缩组 **操作** 列下的 **停用**，然后单击待启用伸缩组 **操作** 列下的 **启用**。

检查伸缩组 **状态** 列下的状态是否正确。

修改伸缩组

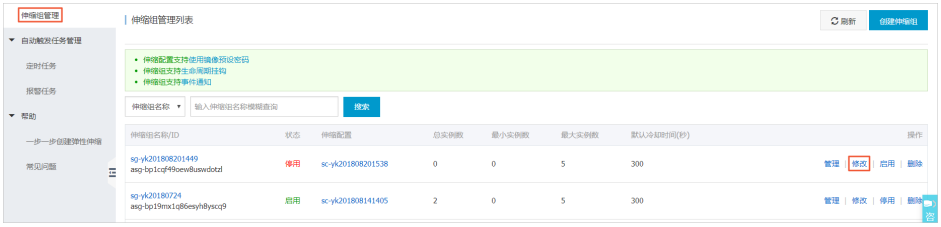
创建伸缩组 后，如果当前伸缩组属性不再满足需求，您可以直接修改一项或者多项属性，无需重新创建。

> **说明**：如果修改 *伸缩最大实例数* 和 *伸缩最小实例数*，导致伸缩组内的 ECS 实例数超过上限或低于下限，弹性伸缩会自动加入或移出 ECS 实例，保证 ECS 实例数量满足要求。

操作步骤

请参照以下步骤修改一个伸缩组的属性：

在 **伸缩组管理** 页面中，单击待修改伸缩组 **操作** 列下的 **修改**。



在 **修改伸缩组** 对话框中，根据需要修改参数，然后单击 **提交**。

说明：有关伸缩组的属性，请参阅 创建伸缩组 下的 **伸缩组属性**。

删除伸缩组

如果不再需要某个伸缩组，您可以将其删除，空出伸缩组配额。

说明：删除伸缩组会同时组内的伸缩配置和伸缩规则。如果组内存在运行中的 ECS 实例，弹性伸缩会先停止 ECS 实例，然后移出所有手动添加的 ECS 实例并释放所有自动创建的实例。

操作步骤

请参照以下步骤删除一个伸缩组：

在 **伸缩组管理** 页面中，单击要删除的伸缩组 **操作** 列下的 **删除**。

在 **删除伸缩组** 对话框中，单击 **确定**。

在 **伸缩组管理** 页面中，单击 **刷新** 确认删除已经完成。

手动伸缩

添加 ECS 实例

本文介绍如何向伸缩组添加一个 ECS 实例，有关详细的 ECS 实例配置，请参阅 创建伸缩配置。您可以通过两种方式添加 ECS 实例：

- 执行伸缩规则自动创建一台或多台 ECS 实例。自动创建的 ECS 实例会自动满足当前伸缩配置，您无

需担心产生规格限制问题。
- 手动添加一台或多台 ECS 实例。手动添加的 ECS 实例配置与当前伸缩配置没有关联。

下面将详细介绍手动添加 ECS 实例的过程。

限制条件

如果您手动添加 ECS 实例，请确保待添加的 ECS 实例满足以下条件：

- 必须与伸缩组处于同一个地域。
- 不能已加入到其它伸缩组中。
- 状态必须是 运行中。
- 可以为经典网络类型或 VPC 类型，但有以下限制：
    - 当伸缩组为经典网络类型时，只能添加经典网络类型的 ECS 实例。
    - 当伸缩组为 VPC 类型时，只能添加同一 VPC 下的 ECS 实例。

同时，请确保伸缩组满足以下条件：

- 状态必须是 启用。
- 不能存在执行中的伸缩活动。

说明：弹性伸缩服务尽力保证足额弹出待添加的 ECS 实例，但是，如果出现云服务器库存不足、待添加的 ECS 实例数超过伸缩组上限等问题，ECS 实例会无法足额弹出。这种情况下，请您检查伸缩组相关配置定位问题。如果无法解决问题，请提交工单寻求帮助。

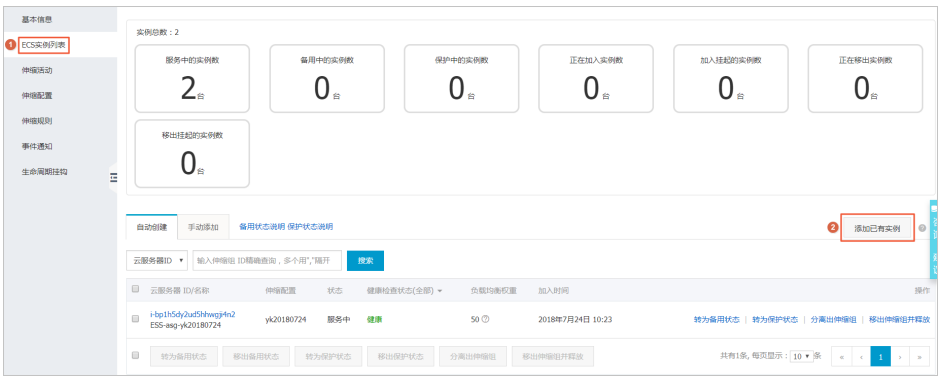
操作步骤

手动添加 ECS 实例可以绕过冷却时间。请参照以下步骤手动添加 ECS 实例：

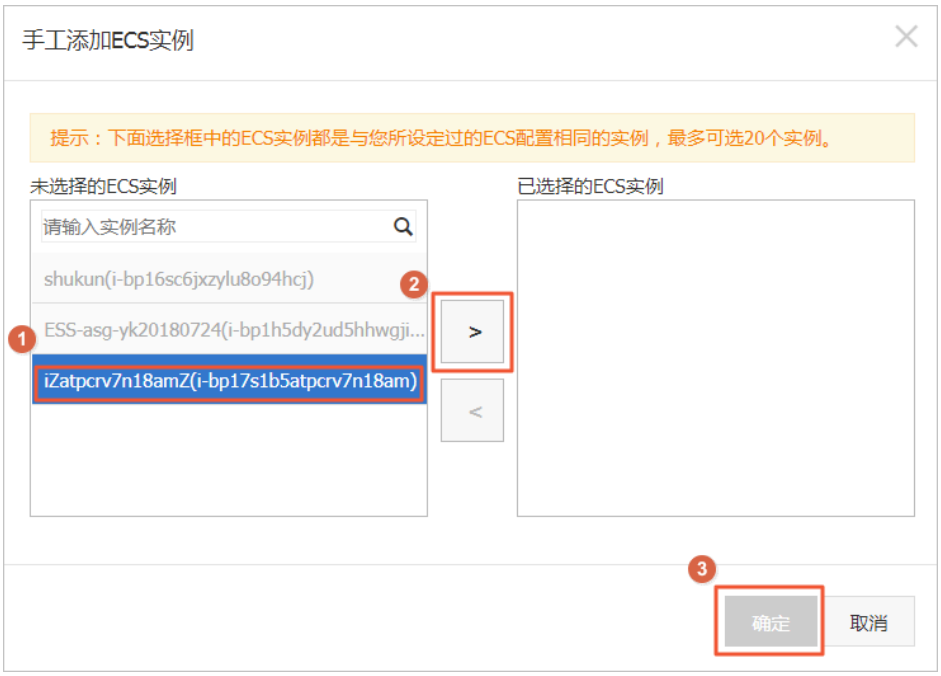
登录弹性伸缩控制台。

在 **伸缩组管理** 页面，单击指定伸缩组 **操作** 列下的 **管理**。

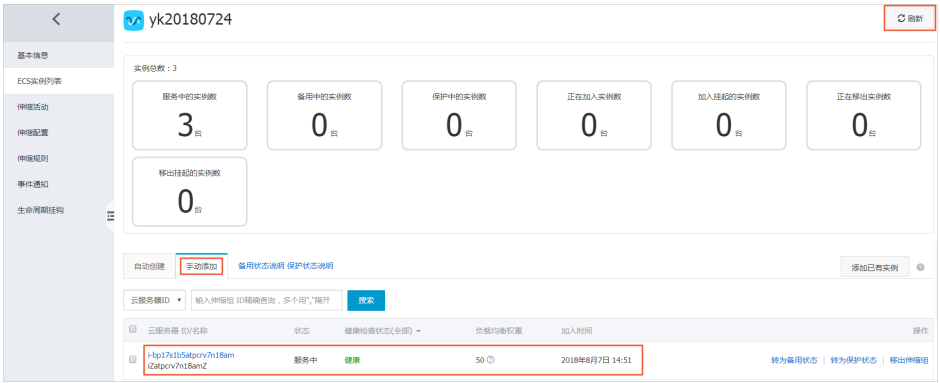
前往 **ECS 实例列表**，单击 **添加已有实例**。



从左侧列表中选择可用的 ECS 实例，单击 > 添加到伸缩组，然后单击 **确定**。



前往 **手动添加** 页签查看结果。



说明：如果页面没有自动刷新，请单击页面右上角的 **刷新** 进行手动刷新。

移出ECS实例

从指定的伸缩组里移出 ECS 实例。

- 从伸缩组移出弹性伸缩自动创建的 ECS 实例时，停止和释放该 ECS 实例。
- 从伸缩组移出用户手工加入的 ECS 实例时，不停止和释放该 ECS 实例。

执行该功能，必须同时满足以下条件：

- 伸缩组为生效（Active）状态。
- 当伸缩组没有伸缩活动正在执行。

当伸缩组没有伸缩活动正在执行时，该功能可以绕过冷却时间（Cooldown）直接执行。

执行功能返回成功，只是表示弹性伸缩服务接受了该功能的调用请求，伸缩活动可以执行，但不代表伸缩活动能够执行成功。用户需要通过返回的 ScalingActivityId 查看该伸缩活动的执行状态。

如果当前伸缩组的实例数（Total Capacity）减去该功能指定的实例数小于 MinSize 时，则执行失败。

示例

定时任务

创建定时任务

如果您可以准确预料业务量的变化时间，就可以提前添加定时任务，在业务高峰到达前准备好足量的计算资源，或者在业务高峰后释放空闲的计算资源。

什么是定时任务

定时任务是在指定时间执行指定伸缩规则的一种预设任务，从而自动扩张或者收缩计算资源，满足业务需求的同时又能控制成本。您还可以为定时任务指定重复周期，通过灵活的规则应对业务量的变化。

> **说明**：一个账号下可以创建的定时任务数量有限，请参阅 使用限制。

由于在同一时刻一个伸缩组中只能存在一个伸缩活动，定时任务还提供了自动重试时间设置功能，避免因为单次执行伸缩规则失败影响定时任务的结果。如果同一分钟内存在多个待执行的定时任务，弹性伸缩会执行最新创建的定时任务。

操作步骤

请参照以下步骤创建一个定时任务：

登录 弹性伸缩控制台。

前往 **自动触发任务管理** 下的 **定时任务** 页面，单击 **创建定时任务**。

在 **创建定时任务** 对话框中，依次指定任务名称、执行时间、伸缩规则、重试过期时间（可选）和重复周期（可选），您还可以添加描述以便日后查看，然后单击 **提交**。

说明：有关定时任务的属性，请参阅 定时任务属性。

定时任务属性

名称	说明	示例
任务名称	名称为2-40个字符，以大小写字母、数字或中文开头，可包含英文句号（.）、下划线（_）或短横线（-）。	st-yk201808301442
描述	描述定时任务的目的、作用等信息，便于日后查看。	月初访问量大，增加 3 台实例。
执行时间	定时任务的触发时间。	2018年9月2日 00:00
伸缩规则	伸缩规则的名称，指定任务触发时执行的伸缩动作。	add3
重试过期时间	时间范围是 0 秒 ~ 21600 秒（6 小时）。如果未能在执行时间触发伸缩动作，弹性伸缩将在重试过期时间内继续执行定时任务。	600
重复周期	重复执行定时任务的周期，您可以选择按日、周、月的方式重复。如果要求多样，也可以使用 Cron 表达式。	按月 每月第 2 天 - 第 3 天执行
重复结束时间	停止重复执行定时任务的时间。	2018年9月30日 00:00

Cron 表达式

Cron 表达式采用 UTC+0 时区，在中国换算成系统本地时间时需要加 8 小时。另外，第一次执行 Cron 表达式的时间必须小于重复结束时间，否则会导致定时任务创建失败。

Cron 表达式是一个字符串，字符串以若干个空格隔开，分为 5 到 7 个域。目前，弹性伸缩定时任务支持 5 域的 Cron 表达式，分别是分钟、小时、日、月和星期，取值范围如下：

域	是否必需	取值范围
分钟	是	[0, 59]
小时	是	[0, 23]
日	是	[1, 31]
月	是	[1, 12]
星期	是	[0, 7]，星期日 = 0 或 7

在同一个域内可以填写多个数值，方法如下：

- 利用逗号（,）指定多个值，例如：“1,3,4,7,8”。
- 利用短横线（-）指定值的范围，例如：“1-6”，结果等同于“1,2,3,4,5,6”。
- 利用星号（*）指定任何可能的值，例如：小时域中填写星号代表每个整点，结果等同于“0-23”。
- 利用正斜线（/）指定间隔频率，例如：小时域中填写“0-23/2”表示每 2 小时执行一次。同时正斜线可以和星号一起使用，例如：小时域中填写“*/3”表示每 3 小时执行一次。

修改、查询、删除定时任务

描述

修改定时任务的属性、查询定时任务的信息、删除一个指定的伸缩规则

报警任务

弹性伸缩报警任务

弹性伸缩（Auto Scaling）报警任务是 弹性伸缩 与 云监控服务（CMS） 深度合作，提供的一种动态管理伸缩组的方式，类似于弹性伸缩定时任务，弹性伸缩报警任务通过触发您指定的伸缩规则来执行伸缩活动，达到调整伸缩组内实例个数的目的。

定时任务可以在您指定的时间执行您指定的伸缩规则，当业务场景在时间上可预料时，能够提前做出响应，但是，在面对突发或者时间上不可预料的业务场景时，定时任务就显得捉襟见肘，此时，就需要报警任务来提供更灵活的触发伸缩规则的方式，在业务高峰期增加伸缩组内实例数量来缓解业务压力，在业务低谷时释放伸缩组内实例，减小生产成本。

报警任务通过监控特定的监控指标，对数据指标进行实时的统计，当统计值满足您指定的报警条件时，触发报警，执行您指定的伸缩规则。使用报警任务，您可以实时的根据业务的变化来不断调整伸缩组内的实例数量，保证您监控的指标维持在您期望的范围内。

弹性伸缩报警任务为您提供了一种通过监控特定监控指标来动态调整伸缩组内实例数量的方法，让您能够根据业务的变化实时的执行指定的伸缩规则，调整伸缩组内的实例数量。

弹性伸缩报警任务升级版

弹性伸缩报警任务已经全新升级，从监控范围、监控方式、监控响应速度等三个方面做出了全方位的优化，升级后的弹性伸缩报警任务将您提供一种更全面，更可靠的利用报警任务动态管理伸缩组的方式。

升级内容主要包括：

- 增加了对系统磁盘，网卡，TCP连接数等监控指标的报警任务支持。
- 报警任务最小统计周期升级到 1分钟，提供更灵敏的监控报警。
- 增加自定义监控，为用户自有监控系统接入弹性伸缩报警任务提供标准化方式。

弹性伸缩报警任务升级版扩展了原有的监控指标，并在原有监控项的基础上，支持用户接入自定义的监控项，提供定制化的报警任务，大大增强了弹性伸缩报警任务的可用性和实用性，满足用户具体的，多样化的需求。

系统监控报警任务

系统监控报警任务的监控指标是云监控为用户采集的 ECS 实例的相关数据指标，以伸缩组作为监控粒度，即以伸缩组内所有实例的监控指标的统计平均值作为伸缩组的指标值。当伸缩组内实例数量发生变化时，监控指标也会同时进行更新。

支持的监控指标

目前，系统监控报警任务支持以下监控指标：

监控指标	单位	适用网络
CPU	%	经典网络、专有网络
内存	%	经典网络、专有网络
系统平均负载	无	经典网络、专有网络
内网出流量	KB/min	经典网络、专有网络
内网入流量	KB/min	经典网络、专有网络
TCP总连接数	个	经典网络、专有网络
TCP已建立连接数	个	经典网络、专有网络
系统盘读BPS	Byte/s	经典网络、专有网络
系统盘写BPS	Byte/s	经典网络、专有网络
系统盘读IOPS	个/s	经典网络、专有网络
系统盘写IOPS	个/s	经典网络、专有网络
内网网卡发包数	个/s	经典网络、专有网络
内网网卡收包数	个/s	经典网络、专有网络
外网出流量	KB/min	经典网络、专有网络
外网入流量	KB/min	经典网络、专有网络
外网网卡发包数	个/s	经典网络
外网网卡收包数	个/s	经典网络

注意事项

- 伸缩组在同一时刻只能执行一个伸缩活动，当伸缩组内存在正在执行的伸缩活动时，由报警任务触发伸缩规则产生的伸缩活动将被拒绝。
- 弹性伸缩报警任务在触发伸缩规则时受伸缩规则冷却时间影响，伸缩规则在冷却时间内时，将拒绝执行伸缩规则。伸缩组内新增加的 ECS 实例从加入伸缩组到完成系统启动配置，部署用户业务，获取到监控数据需要一定的时间（通常需要几分钟），因此您应该根据具体的业务场景，设置合适的冷却时间，防止在新增实例的监控数据缺失的这段时间中，重复触发伸缩规则。

- 弹性伸缩报警任务，默认设置了一分钟的沉默时间，即触发报警之后，一分钟内不会再次触发伸缩规则。
- 部分系统监控项（内存、负载、网卡发包数、TCP 连接数）指标的采集需要为您安装云监控客户端。默认情况下，当您针对需要云监控客户端采集的监控项设置报警任务时，将为报警任务关联的伸缩组内的所有实例安装云监控客户端，同时，将为您在云监控控制台开启 新购ECS自动安装云监控，为您所有新购的 ECS 实例安装云监控客户端。

弹性伸缩自定义监控项报警任务

弹性伸缩自定义监控项报警任务的监控对象为用户自主上报到云监控中的监控指标。在一些场景下，系统监控项可能不包含您所需要的监控指标，您可能拥有自己的一套监控系统，并且关心的是与您特定业务相关的某些指标,自定义监控报警任务，为您自有的监控系统，或者与业务相关的自有监控指标提供了设置报警任务的接入点。

弹性伸缩自定义监控报警任务是针对阿里云云监控服务的自定义监控项设置报警的，用户在使用弹性伸缩自定义监控报警任务之前需要首先向云监控上报自定义监控数据，即自定义监控项。云监控自定义监控是提供给用户自由定义监控项及报警规则的一项服务，通过此服务，用户可以针对自己关心的业务指标进行监控，将采集到监控数据上报至云监控，由云监控来进行数据的处理，并可以对其设置报警规则。

上报监控数据到云监控

云监控的自定义监控服务为您提供了上报监控数据的方式，您可以将自己采集到的时序数据上报到云监控，这样的数据称作时间序列。云监控提供了OpenAPI、Java SDK 和阿里云命令行工具（CLI）三种方式上报数据，这里我们将主要关注如何使用Java SDK的方式上报监控数据。更多详细的信息您可以查看文档 上报监控数据。

使用Java sdk之前您需要首先在项目中引入相应的jar包，如果您使用maven管理项目，您只需要在项目中加入以下依赖：

```
<dependency>
<groupId>com.aliyun</groupId>
<artifactId>aliyun-java-sdk-core</artifactId>
<version>3.2.6</version>
</dependency>
<dependency>
<groupId>com.aliyun.openservices</groupId>
<artifactId>aliyun-cms</artifactId>
<version>0.2.4</version>
</dependency>
```

您可以按照如下的方式向云监控上报自定义监控项：

```
static String endPoint     = "https://metrichub-cms-cn-hangzhou.aliyuncs.com";
CMSClient cmsClient = new CMSClient(endPoint, accAutoScalingKey, accAutoScalingSecret);
CustomMetricUploadRequest request = CustomMetricUploadRequest.builder()
.append(CustomMetric.builder()
.setMetricName("myCustomMetric")//自定义指标名
.setGroupId(54504L)//设置分组id
.setTime(new Date())//时间
.setType(CustomMetric.TYPE_VALUE)//类型为原始值,
.appendValue(MetricAttribute.VALUE, number)//原始值，key只能为这个
.appendDimension("key1", "value1")//添加维度
.appendDimension("key2","value2")
.build())
.build();
CustomMetricUploadResponse response = cmsClient.putCustomMetric(request);//上报
```

在上述的代码片段中，我们上报了一个数据点到云监控。上报时，必须指定groupId参数，即云监控应用分组id，该分组id可以是您在云监控中已创建的应用分组，也可以是一个不存在的应用分组。您可以在 云监控应用分组 创建和查看应用应用分组信息。您上报的自定义监控项（时间序列），可以在云监控控制的 自定义监控 中查看。

我们建议您向一个 云监控中已经存在的应用分组中推送自定义监控数据，云监控的应用分组是针对多种云产品的逻辑分组，向一个已存在的应用分组推送自定义监控数据将使您在需要使用云监控等相关功能时保留扩展的能力。当然，您也可以完全不必理会应用分组的概念，选择向任意一个分组id推送数据。

您上报到云监控的监控数据，云监控会自动帮您进行聚合，当然，当您需要推送的数据量太大时，您也可以选择在本地聚合之后再推送到云监控。详细信息可以参考 上报监控数据。

注意事项

云监控对用户上报监控数据设置了以下限制：

- 单云账号QPS限制为100。
- 单次最多上报100条数据，body最大为256KB。
- “metricName”字段只支持字母、数字、下划线。需要以字母开头，非字母开头会替换为大写“A”，非法字符替换为“_”。
- “dimensions”字段不支持“=”、“&”、“,”，非法字符会被替换为“_”。
- metricName 和dimensions的Key-value 最大均为64字节，超过64字节会被截断。
- 其他限制请关注 计量计费 说明。

创建报警任务

本文介绍如何创建一个报警任务，目前报警任务分为两种类型：系统监控项报警任务和自定义监控项报警任务。

创建系统监控项报警任务

进入 弹性伸缩控制台，分三步创建报警任务：

1. 选中报警任务侧边栏。
2. 点击创建报警任务。
3. 在弹出的创建报警任务对话框中填写定制信息。

示例

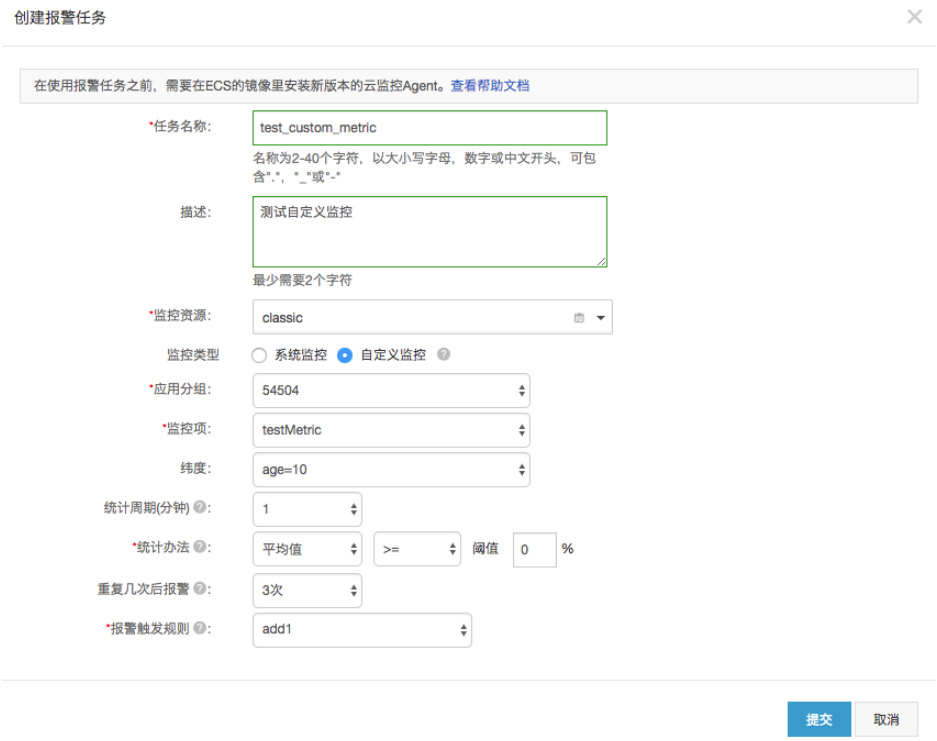
上图中的报警任务信息含义如下：

- test_cpu_alarm 为任务名称，cpu utilization 为任务描述。
- classic 为监控资源，即报警任务监控的伸缩组。
- 系统监控 为监控类型。
- CPU（CPU使用率） 为监控项。
- 1 分钟为统计周期，表示将对每1分钟的监控数据进行统计检查，判断是否满足报警条件。
- 平均值>=50% 为统计方法，重复 3次。这表示，当cpu的使用了在1分钟平均值超过阈值50%，并且连续3次满足统计办法时，将发生报警。
- 伸缩规则add1为报警触发规则，表明在发生报警时，将执行报警规则add1，为伸缩组增加1台机器。

创建自定义监控项报警任务

创建自定义监控项报警任务的过程和创建系统监控类似，唯一的不同在于，系统监控报警任务的监控项是由云监控为用户采集，而自定义监控报警任务的监控项则需要用户自己上报至云监控。

在创建自定义监控类型的报警任务时，首先需要存在已上报的自定义监控项，即时间序列。然后才能针对该时间序列设置报警规则。



在上图创建自定义报警任务之前，已经向云监控推送了一条自定义监控数据流（时间序列）。该时间序列所属的应用分组为54504，监控项名称为testMetric，维度信息为 age=10。

查看修改删除报警任务

您可以在 弹性伸缩控制台 对已创建的报警任务进行查看、修改和删除操作。

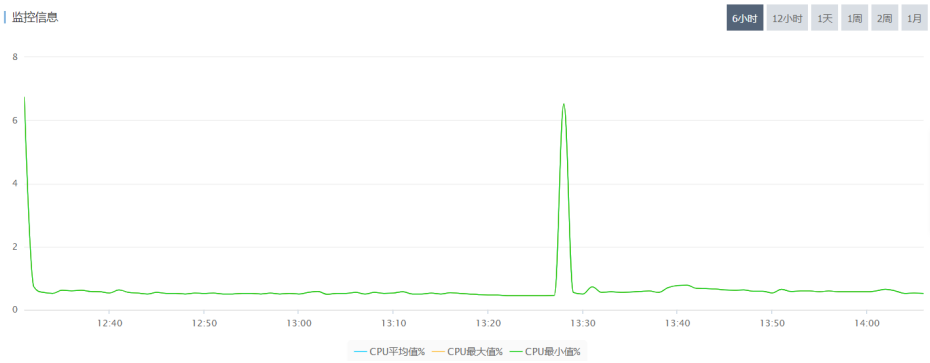
查看监控项详情

在成功创建报警任务之后，我们可以在报警任务列表看到刚刚创建的报警任务。

在系统监控列表，查看创建的系统监控报警任务。

切换到自定义监控列表，查看创建的自定义报警任务。

点击报警任务的名称，将跳转到详情页，在详情页您将看到报警任务对应监控项的历史数据。



修改报警任务

您可以在报警任务的列表页修改报警任务，也可以进入报警任务的详情页修改报警规则。

修改报警任务分为两个部分，一个是修改报警任务的基本信息，另外一个是修改报警规则的触发规则。

修改基本信息包括修改任务名称、监控项、统计周期、统计办法、重复次数等，我们不建议您修改报警任务的监控项，因为，修改监控项意味着监控不同的指标，此时，新建一条针对新指标的报警任务是更合适的方法。

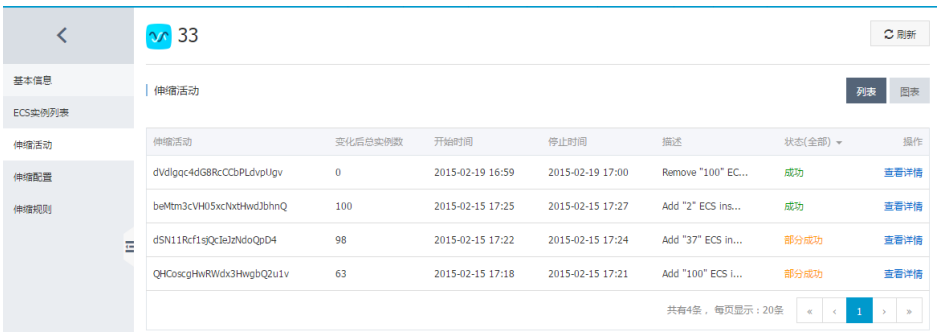
删除报警任务

您可以在报警任务的列表页删除报警任务。

查看伸缩活动

查询伸缩活动的信息，最多只能返回30天之内的伸缩活动。

示例



设置实例备用状态

弹性伸缩支持设置ECS实例备用状态（Standby）。我们不会检查处于备用状态的ECS实例健康状态，亦不会释放ECS实例。

功能特性

设置ECS实例备用状态后：

- ECS实例保持备用状态直至您使其退出备用状态。
- 弹性伸缩不会管理ECS实例的生命周期，而是由您管理。
- 如果对应伸缩组配置了负载均衡，则备用状态ECS实例的负载均衡权重会被置零。
- 您可以 停止 实例、重启 实例或者做其他变配操作，例如 更改ECS实例规格、更换操作系统、初始化磁盘 或者 迁移经典网络到专有网络VPC 等。
- 伸缩组内发生伸缩活动时，不会移除处于备用状态的ECS实例。
- ECS实例被停止或者重启，不会更新ECS实例的健康检查状态。
- 您需要预先从伸缩组内移除该ECS实例才能将其释放。
- 如果您删除伸缩组，ECS实例会自动解除备用状态并随伸缩组一起释放。

解除ECS实例备用状态后：

- ECS实例恢复运行状态。
- 如果伸缩组设置了负载均衡实例，则将负载均衡实例中该ECS实例的权重设置为定义值。
- ECS实例被停止或者重启时，会重新更新健康检查状态。
- 弹性伸缩继续接管ECS实例的生命周期。

控制台操作

设置备用状态

登录 弹性伸缩管理控制台。

选择 **地域**，如华东2。

找到并单击目标伸缩组。

在左侧导航页面，单击 **ECS实例列表**。

勾选目标ECS实例，单击 **转为备用状态**。

解除备用状态

登录 弹性伸缩管理控制台。

选择 **地域**，如华东2。

找到并单击目标伸缩组。

在左侧导航页面，单击 **ECS实例列表**。

勾选目标ECS实例，单击 **移出备用状态**。

API操作

设置备用状态：EnterStandby

解除备用状态：ExitStandby

相关链接

- *云栖社区* 伸缩伸缩支持Standby
- 什么是负载均衡
- 移出不健康ECS实例

查询ECS实例列表

查看 ECS 实例列表

查询伸缩组内 ECS 实例列表。

加入伸缩组的 ECS 实例有两种类型：自动创建的 ECS 实例、手工添加的 ECS 实例。

- **自动创建的 ECS 实例** 是指根据用户的伸缩配置和伸缩规则，由弹性伸缩服务自动创建的 ECS 实例。
- **手工添加的 ECS 实例** 是指不是由弹性伸缩服务创建，但由用户手工添加到伸缩组中的 ECS 实例。

ECS 实例在伸缩组中的生命周期

ECS 实例在伸缩组中的生命周期，通过以下几种状态描述：

- 加入中（Pending） – 表示 ECS 实例正在加入伸缩组，包括创建实例、加入负载均衡、添加 RDS 访问名单等过程。
- 服务中（InService） – 表示 ECS 实例已成功加入伸缩组，并正常提供服务。
- 移出中（Removing） – 表示 ECS 实例正在移出伸缩组。

ECS 实例的健康状态

ECS 实例在伸缩组中的健康状态为：

- 健康（Healthy）
- 不健康（Unhealthy）

本期仅通过 ECS 实例为非“运行中（Running）”状态来判断该实例不健康，弹性伸缩会自动移出伸缩组中不健康的 ECS 实例。对于“自动创建的 ECS 实例”，弹性伸缩会停止和释放该 ECS 实例。对于“手工添加的 ECS 实例”，弹性伸缩不会停止和释放该 ECS 实例。

示例

下图是一个示例。

弹性伸缩服务ESS
伸缩组管理
自动触发任务管理
整体方案创建管理
phpwind-cms-pa...
返回伸缩组管理列表
刷新
基本信息
ECS实例列表
伸缩活动
伸缩配置
伸缩规则
实例总数：2
正在服务中的实例数
2台
正在加入实例数
0台
正在移出实例数
0台
ECS实例列表
自动创建
手动添加
添加已有实例
云服务器名称
伸缩配置
状态
健康检查状态(全部)
加入时间
操作
ESS-sg-phpwind-cms-patch-ecs-i-25pxu6uxw
test
服务中
健康
2014-10-07 17:11
移出伸缩组并释放
ESS-sg-phpwind-cms-patch-ecs-i-25dkmmajr
test
服务中
健康
2014-10-05 17:11
移出伸缩组并释放
移出伸缩组并释放
共有2条，每页显示：10条
1